**Central European Journal of Economic Modelling and Econometrics**

# Bayesian Model Selection in the Analysis of Cointegration

Justyna Wróblewska*

Submitted: 31.08.2008, Accepted: 22.03.2009

**Abstract**

In this paper we present the Bayesian model selection procedure within the class of cointegrated processes. In order to make inference about the cointegration space we use the class of Matrix Angular Central Gaussian distributions. To carry out posterior simulations we use an alorithm based on the collapsed Gibbs sampler. The presented methods are applied to the analysis of the price - wage mechanism in the Polish economy.

**Key Words:** cointegration; Bayesian analysis; Grassmann manifold; Stiefel manifold; posterior probability

**JEL Classification**: C11; C32; C52

*Cracow University of Economics, e-mail: ewroble@cyf-kr.edu.pl

# 1 Introduction

The general aim of this paper is to formally compare competing Bayesian models with the error correction mechanism, which may differ in the lag length, the type of deterministic processes, the number of cointegrating relations and overidentifying restrictions imposed on the cointegration space (e.g. the (trend) stationarity of a given component of the analysed multivariate process) and/or the space of adjustment coefficients (e.g. weak exogeneity of a subset of the analysed variables). The obtained posterior probabilities of each model may be used in further analyses and decision making processes.
The methods presented here will be applied to the analysis of the price - wage mechanism in the Polish economy. The analysed 52 quarterly data include five variables: average wages, current prices (W), price index of consumer goods (P), labour productivity, constant prices (Z), price index of imported goods (M), rate of unemployment (U) and covers the thirteen year period ranging from 1995 to 2007.
This paper is laid out as follows: section 2 introduces the sample model, sections 3 and 4 present the set of competing models and some aspects of the Bayesian model selection respectively, section 5 contains an empirical example and section 6 concludes.

# 2 The basic Bayesian Vector Error Correction Model

Let us consider the $n$-dimensional cointegrated process $\{x_t\}_{t=1,2,\ldots,T}$, where $x_t = (x_{t1}, x_{t2}, \ldots, x_{tn})'$, $t = 1, 2, \ldots, T$. According to the Granger representation theorem any cointegrated process may by written in the error correction form (Strachan, van Dijk 2007):

$$\begin{aligned} \Delta x_t &= \alpha(\beta^{+\prime} x_{t-1} + \varphi_1' d_{1t}) + \Gamma_0 w_t + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \varphi_2 d_{2t} + \varepsilon_t = \\ &= \alpha\beta' z_{1t} + \Gamma' z_{2t} + \varepsilon_t \end{aligned} \tag{1}$$

and in matrix notation

$$Z_0 = Z_1\beta\alpha' + Z_2\Gamma + E = Z_1\Pi' + Z_2\Gamma + E \tag{2}$$

where $Z_0 = (\Delta x_1, \Delta x_2, \ldots, \Delta x_T)'$, $Z_1 = (z_{11}, z_{12}, \ldots, z_{1T})'$, $z_{1t}' = (x_{t-1}'\, d_{1t}')$, $Z_2 = (z_{21}, z_{22}, \ldots, z_{2T})'$, $z_{2t}' = (w_t', \Delta x_{t-1}', \Delta x_{t-2}', \ldots, \Delta x_{t-k+1}', d_{2t}')$, $\beta = (\beta^{+\prime}\, \varphi_1')'$, $\Gamma = (\Gamma_0, \Gamma_1, \ldots, \Gamma_{k-1}, \varphi_2)'$, $E = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)'$, $\varepsilon_t \sim iiN^n(0, \Sigma), t = 1, 2, \ldots, T$. Matrices $d_{1t}$, $d_{2t}$ introduce deterministic trends to the VECM form and $w_t$ contains other non-random regressors, $\alpha$ is the $n \times r$ matrix of adjustment coefficients, $\beta$ is the $m \times r$ matrix containing cointegrating vectors; $m \geq n$ and $m = n$ if there are no deterministic components in the cointegrating relations. Both matrices, $\alpha$ and $\beta$, are of rank $r$, where $0 \leq r \leq n$. For $r = n$ we assume $\alpha = I_n$.

The decomposition of the reduced rank $\Pi$ matrix as the product of the two full column rank matrices $\alpha$ and $\beta$ is very convenient because of the natural interpretation, but its main drawback is ambiguity, i.e. for any $r \times r$ full rank matrix $C : \Pi = \alpha\beta' = \alpha CC^{-1}\beta'$, in other words - data contain information only about the cointegration space. The cointegration space is the element of the Grassmann manifold ($G_{r,m-r}$), which is the set of all $r$ dimensional subspaces in $\mathbb{R}^m$. Following Strachan and Inder (2004) approach we make use of the many-to-one relation between the Stiefel manifold ($V_{r,m}$) and the Grassmann manifold. The points of the Stiefel manifold are $r-$frames, i.e. sets of $r$ orthonormal vectors in $\mathbb{R}^m$. In order to make Bayesian inference about the cointegration space, it is possible to work with $\beta \in V_{r,m}$ and adjust the integrals with respect to $\beta$ by the volume of the orthogonal group $O(r)$ to account for the fact that the dimension of $V_{r,m}$ exceeds the dimension of $G_{r,m-r}$. Assuming that $\beta \in V_{r,m}$ we do not exclude any direction of the cointegration space, and by imposing uniform prior on the set of semi-orthogonal matrices $V_{r,m}$, we impose uniform prior on the Grassmann manifold ($G_{r,m-r}$). It is worth to emphasize that both manifolds considered here are compact, so uninformative priors are proper and unique, so, even if we want to be uninformative about the cointegrating space, we can calculate posterior odds ratios in order to choose between competing models (see e.g. James 1954 for the formal discussion and Strachan, Inder 2004 for the discussion in the context of VECM). Strachan and Inder (2004) present the construction of not only uninformative, but also informative priors in which the researcher can incorporate prior knowledge about the cointegration space. They use the class of Matrix Angular Central Gaussian distributions, $MACG$ (Chikuse 1990). To carry out efficient posterior simulations in such models, Koop, León-González, Strachan (2006) (see also Koop, Potter, Strachan 2008) developed two algorithms, based on a collapsed Gibbs sampler (see e.g. Liu 1994). We present that one of their algorithms, which can be used when matrices $\beta$ and $\alpha$ can have different dimensions and so can be used in models with deterministic terms in cointegrating relations and/or in over-identified models. In this algorithm the following parameterisations of the $\Pi$ matrix will be used:

$$\tilde{\alpha}\tilde{\beta}' = \tilde{\alpha}DD^{-1}\tilde{\beta}' = (\tilde{\alpha}D)(\tilde{\beta}D^{-1})' \equiv \alpha\beta', \tag{3}$$

where $D$ is an unidentified $r \times r$ symmetric positive definite matrix. In the first parametrisation both matrices $\tilde{\beta}$ and $\tilde{\alpha}$ have real space as their support. For $D = (\tilde{\beta}'\tilde{\beta})^{\frac{1}{2}}$, $\beta$ in the second parametrisation has orthonormal columns and $\alpha$ remains unrestricted. The sampler alternates between these parameterisations and involves draws from the Normal and inverted Wishart distributions (also from the inverted Gamma distributions, when we treat scalars controlling the precisions of prior distributions as additional parameters).

We impose on $\tilde{\beta}$ and $\tilde{\alpha}$ matrix Normal distributions:

$$\tilde{\beta}|\tau, r, m \sim mN_{m\times r}(0, m^{-1}I_r, P_\tau) \tag{4}$$

and

$$\tilde{\alpha}|\Sigma, \nu, r \sim mN_{n\times r}(0, \nu I_r, \Sigma) \quad (5)$$

Parameters $\nu$ and $\tau$ control degrees of informativeness of the distributions stated above and we impose for them inverted Gamma distributions $\nu \sim iG(s_\nu, n_\nu)$, $\tau \sim iG(s_\tau, n_\tau)$. In the $P_\tau$ matrix the researcher may incorporate prior knowledge. In order to obtain the prior distribution for $\beta$, we use theorem 1.

**Theorem 1 (Chikuse, 1990, 2003)** *Let us suppose that $Z$ has the $m \times r$ matrix-variate central Normal distribution with parameter $\Omega$ ($Z \sim mN_{m\times r}(0, I_r, \Omega)$), whose density is*

$$p(Z) = (2\pi)^{-\frac{rm}{2}} |\Omega|^{-\frac{r}{2}} exp\left[-\frac{1}{2} tr(Z'\Omega^{-1}Z)\right],$$

*where $\Omega$ is $m \times m$ positive definite matrix. Then we have the density of its orientation $H_Z = Z(Z'Z)^{-\frac{1}{2}}$*

$$p(H_Z) = |\Omega|^{-\frac{r}{2}} |H_Z'\Omega^{-1}H_Z|^{-\frac{m}{2}}. \quad (6)$$

From the uniqueness of the polar decomposition (see e.g. Cadet 1996) of $\tilde{\beta}$ ($\beta = H_{\tilde{\beta}}$) and according to Theorem 1, $\beta$ has matrix angular central Gaussian distribution with parameter $P_\tau$: $\beta|\tau, r \sim MACG(P_\tau)$. If we assume that $P_\tau = I_m$ we get a uniform distribution over the Stiefel manifold and so a uniform distribution over the Grassmann manifold.
The priors for the remaining parameters are:

- inverted Wishart for $\Sigma$: $\Sigma \sim iW(S, q)$ (we opt for the informative prior for $\Sigma$, because, in order to estimate the marginal likelihood of the data, we will use the Newton - Raftery method),
- matrix Normal for $\Gamma$: $\Gamma|\Sigma, h \sim mN(0, \Sigma, hI)$,
- inverted Gamma for $h$, if the researcher wants it to be estimated: $h \sim iG(s_h, n_h)$.

# 3 The set of competing models

The considered VECM forms may differ in the lag length of VAR ($k$), the type of deterministic processes ($d$), the rank of the $\Pi$ matrix ($r$), the structural overidentifying restrictions imposed on the cointegrating space ($o$) and the space spanned by the matrix of adjustment coefficients ($e$).
We will consider five commonly used forms of the deterministic trends in the VECM form (see e.g. Johansen 1996, Strachan, van Dijk 2007).

**d=1:** $\varphi_1 d_{1t} = \mu_1 + \delta_1 t$ and $\varphi_2 d_{2t} = \mu_2 + \delta_2 t$, there is a linear trend in the cointegrating relations and the process $x_t$ has a quadratic trend,

**d=2:** $\varphi_1 d_{1t} = \mu_1 + \delta_1 t$ and $\varphi_2 d_{2t} = \mu_2$, there is still a linear trend in the cointegrating relations and a linear trend in the process $x_t$,

**d=3:** $\varphi_1 d_{1t} = \mu_1$ and $\varphi_2 d_{2t} = \mu_2$, the model allows for a linear trend in the process $x_t$ and a non-zero mean in the cointegrating relations,

**d=4:** $\varphi_1 d_{1t} = \mu_1$ and $\varphi_2 d_{2t} = 0$, there is no trend in the process $x_t$ and a non-zero mean in the cointegrating relations,

**d=5:** $\varphi_1 d_{1t} = 0$ and $\varphi_2 d_{2t} = 0$, there is no trend in the components of the process $x_t$ and cointegrating relations have zero mean.

The most commonly tested structural restrictions imposed on the cointegration space are of the form:

**1.** $sp(\beta) \subseteq sp(H)$ i.e. $\beta = H\psi$, where $H_{m\times s}$ is a known matrix and $\psi_{s\times r}\,(r \leq s < m)$ contains unknown parameters. This restriction is imposed on all cointegrating vectors. The cointegration space is fully determined. As $sp(H) = sp(H(H'H)^{-1/2})$ we can assume that $H$ is an element of the Stiefel manifold ($H \in V_{s,m}$). This restriction leads to the model of the following form:

$$Z_0 = (Z_1 H)\psi\alpha' + Z_2\Gamma + E = (Z_1 H)\tilde{\psi}\tilde{\alpha}' + Z_2\Gamma + E. \tag{7}$$

We know that $\beta'\beta = I_r$, so $I_r = \beta'\beta = \psi' H' H\psi = \psi'\psi \Rightarrow \psi \in V_{r,s}$ and we can, for example, impose for this matrix $MACG(\tilde{P}_\tau)$ distribution (i.e. $mN(0, s^{-1}I_r, \tilde{P}_\tau)$ for $\tilde{\psi}$).

**2.** $sp(b) \subseteq sp(\beta)$ i.e. $\beta = (b \quad \phi) = (b \quad b_\perp\psi)$, where $b_{m\times s}\,(s \leq r)$ is a matrix containing known cointegrating vectors and $\phi_{m\times(r-s)}$ contains unknown cointegrating vectors. As $sp(b) = sp(b(b'b)^{-1/2})$ we can assume that $b$ is an element of the Stiefel manifold ($b \in V_{r-s,m}$). According to Theorem 2 applied to $X_1 = b$, $G(X_1) = b_\perp$ and $U_1 = \psi$, inference in such a model can be made as in the basic model:

$$Z_0 = (Z_1 b_\perp)\psi\alpha_2' + (Z_1 b)\alpha_1' + Z_2\Gamma + E = (Z_1 b_\perp)\tilde{\psi}\tilde{\alpha_2}' + (Z_1 b)\alpha_1' + Z_2\Gamma + E,$$

$$\tilde{\psi}|\tau, r-s, m-s \sim mN_{m-s\times r-s}(0, (m-s)^{-1}I_{r-s}, \tilde{P}_\tau), \qquad \psi|\tau \sim MACG(\tilde{P}_\tau),$$

$$\alpha_1|\nu, \Sigma \sim mN(0, \nu I_s, \Sigma), \qquad \tilde{\alpha_2}|\nu, \Sigma \sim mN(0, \nu I_{r-s}, \Sigma)$$

**Theorem 2 (Chikuse, 1990, 2003)** *Let us write a random matrix $X$ on $V_{r,m}$ as $X = (X_1\, X_2)$ with $X_1$ and $X_2$ being $m\times s$ and $m\times(r-s)$ matrices, respectively ($0 < s < r$). Then we can write $X_2 = G(X_1)U_1$, where $G(X_1)$ is any matrix chosen so that $(X_1\, G(X_1))$ is orthogonal, and as $X_2$ runs over $V_{r-s,m}$, $U_1$ runs over $V_{r-s,m-s}$ and the relationship is one-to-one. The differential form $[dX]$ for the normalized invariant measure on $V_{r,m}$ is decomposed as the product $[dX] = [dX_1][dU_1]$ of the forms $[dX_1]$ and $[dU_1]$ on $V_{s,m}$ and $V_{r-s,m-s}$, respectively.*

This type of restriction can be used to check whether one of the components of the process $x_t$ is stationary. By including in the model different types of deterministic trend, the hypothesis about trend stationarity can also be tested.

**3.** $\beta = (H_1\psi_1,\, H_2\psi_2, \ldots, H_l\psi_l)$**,** where $H_i, \quad i = 1, 2, \ldots, l$ are $m \times s_i$ matrices, $\psi_i$ are $s_i \times r_i$ matrices, $r_i \leq s_i, \quad l \leq r, \sum_{i=1}^{l} r_i = r$. This type of restriction contains the forms stated above as special cases

The restrictions for the adjustments coefficients can be formulated in a similar way to those imposed on $\beta$. One of the most important restrictions is the sufficient condition for weak exogeneity (for the estimation of the long-run parameters and the remaining adjustment parameters), i.e. $\alpha_2 = 0$, where $\alpha_2$ contains adjustment parameters for the components of $x_t$, say $x_{2t}$, which weak exogeneity we check (see e.g. Urbain 1992). This hypothesis can be formulated as $sp(\alpha) \subseteq sp(\tilde{A})$, i.e $\alpha = \tilde{A}\psi$, where $\tilde{A} = (I_s \quad 0)$. As a prior distribution for $\tilde{\psi} = \psi(\tilde{\beta}'\tilde{\beta})^{-\frac{1}{2}}$ we could use $mN(0, \nu I_r, \tilde{A}'\Sigma\tilde{A})$.

# 4 Bayesian model selection

Bayesian methodology enables us to compare different models through the posterior probability (see e.g. Zellner 1971, Osiewalski 2001, Pajor 2003). The model with the highest posterior probability is usually considered the best.
Let us consider a set of non-nested competing Bayesian models $\{M_\xi\colon \quad \xi = (k, d, r, o, e) \in \Xi\}$:

$$M_\xi\colon p_\xi(x, \theta_{(\xi)}) = p_\xi(\theta_{(\xi)}) p_\xi(x|\theta_{(\xi)}), \qquad \xi \in \Xi,$$

where $x$ denotes the data, $\theta_{(\xi)} \in \Theta_{(\xi)}$ is the vector of parameters of the $M_\xi$ model and $p_\xi(\theta_{(\xi)})$ is the prior density. Using the Bayes theorem, we can evaluate the posterior probability of each model:

$$p(M_\xi|x) = \frac{p(M_\xi)p(x|M_\xi)}{\sum_{\zeta \in \Xi} p(M_\zeta)p(x|M_\zeta)}, \tag{8}$$

where

$$p(x|M_\xi) = \int_{\Theta_{(\xi)}} p_\xi(x|\theta_{(\xi)}) p_\xi(\theta_{(\xi)}) \mathrm{d}\theta_{(\xi)}, \qquad \xi \in \Xi$$

is the marginal density of the data under model $M_\xi$. In this paper we will evaluate these integrals using the Newton-Raftery (N-R) method within the collapsed Gibbs sampler, so as the estimate for $p(x|M_\xi)$ the harmonic mean of the likelihood values of a sample from the posterior distribution will be used (Newton, Raftery 1994, Kass, Raftery 1995):

$$\hat{p}(x|M_\xi) = \left( \frac{1}{m} \sum_{i=1}^{m} p_\xi(x|\theta_{(\xi)i})^{-1} \right)^{-1}. \tag{9}$$

The N-R estimator $\hat{p}(x|M_\xi)$ converges almost surely to the true value $p(x|M_\xi)$ as $m \to \infty$, but it does not have a finite variance. The prior probabilities of competing models, $p(M_\xi)$, $\xi \in \Xi$, are assumed by the researcher. Often we want to treat all models as equally probable. Assuming equal prior probabilities for elements of the set of different VAR/VECM forms is not straightforward, because some of combinations of the individual elements of $\xi$ are impossible (e.g. $r = 0$ and $o \neq 0$) or observationally equivalent to one another (e.g $r = 0, d = 2$ and $r = 0, d = 3$). To overcome this difficulty we will use an algorithm proposed by Strachan and van Dijk (2007):

1. Assume equal probabilities for individual elements of $\xi$, e.g. for $r \in \{0, 1, \dots, n\}$ impose $p(r) = (n + 1)^{-1}$.
2. For all combinations of the individual elements of $\Xi$ set weights: $k(M_\xi) = p(k)p(d)p(r)p(o)p(e)$.
3. Set $k(M_\xi) = 0$ for all impossible combinations.
4. Set $k(M_\xi) = 0$ for all but one observationally equivalent combinations.
5. Compute prior model probabilities as

$$p(M_\xi) = \frac{k(M_\xi)}{\sum_{\zeta \in \Xi} k(M_\zeta)}.$$

# 5 An empirical example: the analysis of the price inflation in Poland

The methods presented above will be used in the analysis of the price - wage spiral in the Polish economy. The classical analysis (using the Johansen procedure) of the price - wage mechanism in the Polish economy is presented by: Welfe, Majsterek, Florczak (1994), Welfe, Majsterek (2000, 2002), and in the Bayesian approach by Wróblewska (2008). Short-run analysis of price inflation in Poland from a Bayesian perspective is presented by Osiewalski, Welfe (1998). The analysed 52 quarterly data include five variables: average wages, current prices (W), price index of consumer goods (P), labour productivity, constant prices (Z), price index of imported goods (M), rate of unemployment (U) and covers the thirteen year period ranging from 1995 to 2007.
Figure 1 (see the Appendix) presents plots of the analysed data in levels and in first differences. Small letters denote natural logarithms of the original variables.
These graphs suggest that we can not assume constant variance for all but productivity, which seems to be trend stationary. As proposed by Johansen (1996, p. 74) we will check this inside the VECM form by testing the hypothesis $sp(b) \subseteq sp(\beta)$, i.e. $\beta = (b,\ \phi)$, where e.g. for the third variable in the five-dimensional process $b' = (0, 0, 1, 0, 0)$.
The visual inspection of the analysed variables may also suggest that wages and prices

are trend stationary if we incorporate quadratic trends for the variables in the model. However their graphs look even too smooth and for this reason we should analyse theirs stochastic nature with more care. When we look at the plots of their first differences we do not observe any substantial mean reversion. These features may suggest that wages and prices are integrated of order two. If this is true the data may give higher posterior probability for models with the $\Pi$ matrix of low rank (close to zero).
We will consider the set of models, which differ in the number of lags $k \in \{1, 2, 3\}$, deterministic terms $d \in \{1, 2, 3, 4, 5\}$, the number of stable equilibrium relations $r \in \{0, 1, 2, 3, 4, 5\}$, the overidentifying restrictions on these relations $o \in \{0, 1, 2\}$ and one restriction for the adjustments coefficients $e \in \{0, 1\}$. Table 1 presents all considered overidentifying restrictions. Homogeneity condition is an example of the restriction imposed on all cointegrating vectors $sp(\beta) \subseteq sp(H)$,i.e. $\beta = H\varphi$, where

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

There is 540 $(k, d, r, o, e)$ combinations. After removing all impossible combinations and all but one equivalent combinations we are left with 366 different models. We wish to treat them as equally possible so we compute the prior model probabilities in the way suggested by Strachan and van Dijk (2007), which gives $p(M_{(k,d,r,o,e)}) = 0.0027$. We analyse quarterly data, so we have decided to incorporate zero mean seasonal dummies in the model. The results are based on the following priors:

- $\tilde{\beta}|r, m \sim mN(0, m^{-1}I_r, I_m)$, which leads to $\beta|r \sim MACG(I_m)$,
- $\tilde{\alpha}|\nu, r, \Sigma \sim mN(0, \nu I_r, \Sigma)$,
- $\Sigma \sim iW(I_n, 7)$,
- $\Gamma|\Sigma, h \sim mN(0, \Sigma, hI)$,
- $\nu \sim iG(2, 3)$ $(E(\nu) = 1,\ Var(\nu) = 1)$,
- $h \sim iG(0.02, 3)$ $(E(h) = 0.01,\ Var(h) = 0.0001)$
- $p(M_{(k,d,r,o,e)}) = 0.0027$.

The joint prior resulting from this specification has been truncated by the stability condition imposed on the parameters of the cointegrated process.
Table 2 presents the most probable models (with posterior probability not less than assumed prior probability $p(M_{(k,d,r,o,e)})$). Their posterior probabilities sum up to 0.992. The results presented in Table 2 confirm our hypothesis about trend stationarity of

productivity. The data give much support for the model with only one cointegrating relation, which may be a signal that the data are $I(2)$. We have decided to check the characteristic roots for the models listed in Table 2. The posterior means and standard deviations (in brackets) of the moduli of these characteristic roots are given in Table 3. In all cases apart from unit roots we have additional roots higher than 0.9

Table 1: The overidentifying restrictions

| | |
|---|---|
| $o = 0$ | no restriction for the cointegration space |
| $o = 1$ | trend-stationarity of labour productivity |
| $o = 2$ | homogeneity condition |
| $e = 0$ | no restriction for the adjustment coefficients |
| $e = 1$ | weak exogeneity of import prices |

(very close to 1). This seems to confirm our presumptions that there is $I(2)$ unit root in the data. For these reasons we have decided to compare the same set of models for the transformed data. We have replaced prices (both domestic and import) with inflation, and wages with their first differences. Using such transformation, we loose information about some of the long-run properties of the analysed data. We can not use the nominal-to-real transformation, because the data do not give much support for the long-run price homogeneity (the posterior probability of this restriction was 0.001). Table 4 presents the most probable models for the transformed data (with

Table 2: The most probable models

| $k$ | $d$ | $r$ | $o$ | $e$ | $p(M_{(k,d,r,o,e)}\|x)$ | $log_{10}(\hat{p}(x\|M_{(k,d,r,o,e)}))$ |
|---|---|---|---|---|---|---|
| 2 | 3 | 1 | 1 | 0 | 0.615 | 52.068 |
| 2 | 2 | 2 | 1 | 0 | 0.229 | 51.639 |
| 2 | 3 | 2 | 1 | 1 | 0.077 | 51.164 |
| 2 | 3 | 2 | 1 | 0 | 0.071 | 51.131 |

posterior probability not less than assumed prior probability $p(M_{(k,d,r,o,e)})$). Their posterior probabilities sum up to 0.9972. The model with only one cointegrating relation with assumed stationarity of productivity is still in the group of the most probable models, but it achievied lower probability than the model with two cointegrating relations and the same restriction imposed on the cointegrating space. These

Table 3: The posterior means and standard deviations (in brackets) of the moduli of the characteristic roots of the most probable models

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $M_{(2,3,1,1,0)}$ | 1 | 1 | 1 | 1 | 0.989 | 0.573 | 0.214 | 0.160 | 0.113 | 0.065 |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0093) | (0.1442) | (0.0892) | (0.0674) | (0.0558) | (0.0474) |
| $M_{(2,2,2,1,0)}$ | 1 | 1 | 1 | 0.988 | 0.945 | 0.628 | 0.221 | 0.166 | 0.118 | 0.067 |
| | (0.0000) | (0.0000) | (0.0000) | (0.0105) | (0.0542) | (0.1646) | (0.0913) | (0.0689) | (0.0572) | (0.0488) |
| $M_{(2,3,2,1,1)}$ | 1 | 1 | 1 | 0.989 | 0.945 | 0.620 | 0.207 | 0.155 | 0.110 | 0.063 |
| | (0.0000) | (0.0000) | (0.0000) | (0.0101) | (0.0543) | (0.1708) | (0.0872) | (0.0658) | (0.0544) | (0.0461) |
| $M_{(2,3,2,1,0)}$ | 1 | 1 | 1 | 0.988 | 0.941 | 0.633 | 0.220 | 0.165 | 0.117 | 0.067 |
| | (0.0000) | (0.0000) | (0.0000) | (0.0107) | (0.0559) | (0.1636) | (0.0911) | (0.0686) | (0.0569) | (0.0486) |

results confirm our presumtions about the nature of the analysed process and show that the price - wage mechanism in the Polish economy should be rather analysed in the framework of models which take into account the possibility of $I(2)$ nature of the nominal variables (see e.g. Johansen 1996, Banerjee, Cockerell, Russell 2001, Kongsted 2003, in the context of the relationship of prices and wages in the Polish economy see Kelm, Majsterek 2007 and for a Bayesian perspective see Strachan 2007).

Table 4: The most probable models for the transformed data

| $k$ | $d$ | $r$ | $o$ | $e$ | $p(M_{(k,d,r,o,e)}\|x)$ | $log_{10}(\hat{p}(x\|M_{(k,d,r,o,e)}))$ |
|---|---|---|---|---|---|---|
| 2 | 2 | 2 | 1 | 0 | 0.735 | 49.221 |
| 2 | 3 | 1 | 1 | 0 | 0.206 | 48.669 |
| 2 | 2 | 2 | 1 | 1 | 0.021 | 47.684 |
| 2 | 4 | 2 | 2 | 1 | 0.013 | 47.467 |
| 2 | 5 | 2 | 2 | 1 | 0.0093 | 47.325 |
| 2 | 3 | 2 | 2 | 1 | 0.0089 | 47.304 |
| 2 | 4 | 2 | 2 | 0 | 0.004 | 46.946 |

# 6 Conclusions

The methods of comparing VECM representaions presented here can give us much insight into the nature of the analysed multivariate processes. It should be emphasized that in the Bayesian framework it is possible to formaly compare even a very large and complex set of competing models and thus to avoid problems occuring in sequential testing procedures within the traditional (non-Bayesian) approach to cointegration.
The analysis of the price-wage spiral in the Polish economy showed that some of the variables may be integrated of order two. The complete Bayesian I(2) analysis in that case would be both statisticaly and empirically interesting and important.

# Acknowledgements

I would like to thank Robert Kelm for the data. Useful comments by two referees are gratefully acknowledged.

# References

[1] Banerjee A., Cockerell L., Russell B., (2001), An $I(2)$ Analysis of Inflation and the Markup, *Journal of Applied Econometrics* 16, 221-240.

[2] Cadet A., (1996), Polar Coordinates in the $\mathbb{R}^{np}$; Application to the Computation of the Wishart and Beta Laws, *Sankhya: The Indian Journal of Statistics* 58, 101-114.

[3] Chikuse Y., (1990), The Matrix Angular Central Gaussian Distribution, *Journal of Multivariate Analysis* 33, 265-274.

[4] Chikuse Y., (2003), Statistics on Special Manifolds, *Lecture Notes in Statistics*, vol. 174, Springer-Verlag, New York.

[5] James A.T., (1954), Normal Multivariate Analysis and the Orthogonal Group, *Annals of Mathematical Statistics* 25, 40-75.

[6] Johansen S., (1996), *Likelihood-based Inference in Cointegrated Vector Auto-Regressive Models*, New York: Oxford University Press.

[7] Kass R.E., Raftery A.E., (1995), Bayes Factors, *Journal of the American Statistical Association* 90, 773-795.

[8] Kelm R., Majsterek M., (2005), An $I(2)$ Analysis of Inflationary Processes in Poland, [in:] *Macromodels 2004, Problems of Building and Estimation of Econometric Models, "Acta Universitatis Lodziensis", Folia Oeconomica* 190, [ed.:] W. Welfe, A. Welfe, Łódź, 55-72.

[9] Kelm R., Majsterek M., (2006), The $I(2)$ Analysis of Money Demand and Inflation in Poland in the Transition Period 1995 - 2005, [in:] , *32-th International Conference Macromodels'2005*, [ed.:] W. Welfe, A. Welfe, Łódź, 49-72.

[10] Kongsted H. Ch., (2003), An $I(2)$ Cointegration Analysis of Small-Country Import Price Determination, *Econometrics Journal* 6, 53-71.

[11] Koop G., León-González R., Strachan R., (2006), Efficient Posterior Simulation for Cointegrated Models with Priors on the Cointegration Space, Working Paper No. 05/13, University of Leicester, Department of Economics.

[12] Koop G., Potter S. M., Strachan R., (2008), Re-examining the Consumption-Wealth Relationship: the Role of Model Uncertainty, *Journal of Money, Credit and Banking* 40, 341-367.

[13] Liu J.S., (1994), The Collapsed Gibbs Sampler in Bayesian Computations with Applications to a Gene Regulation Problem, *Journal of the American Statistical Association* 89, 958-965.

[14] Newton M.A., Raftery A.E., (1994), Approximate Bayesian Inference by the Weighted Likelihood Bootstrap (with discussion), *Journal of the Royal Statistical Society*, Ser. B 56, 3-48.

[15] Osiewalski J., (2001), *Ekonometria bayesowska w zastosowaniach* [Bayesian Econometrics in Applications], Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków. (in Polish)

[16] Osiewalski J., Welfe A., (1998), The Price-Wage Mechanism: An Endogenus Switching Model, *European Economic Review* 42, 365-374.

[17] Pajor A., (2003), *Procesy zmienności stochastycznej SV w bayesowskiej analizie finansowych szeregów czasowych* [Stochastic Volatility (SV) Processes in Bayesian Analysis of Financial Time Series], Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków. (in Polish)

[18] Strachan R.W., (2007), Bayesian Inference in Cointegrated I(2) Systems: a Generalisation of the Triangular Model, *Econometric Reviews* 26, 439-468.

[19] Strachan R.W., Inder B., (2004), Bayesian Analysis of the Error Correction Model, *Journal of Econometrics* 123, 307-325.

[20] Strachan R.W., van Dijk H.K., (2007), Bayesian Model Averaging in Vector Autoregressive Processes with an Investigation of Stability of the US Great Ratios and Risk of a Liquidity Trap in the USA, UK and Japan, EI 2007-11, Econometric Institute Report, Erasmus University Rotterdam.

[21] Urbain J.P., (1992), On Weak Exogeneity in Error Correction Models, *Oxford Bulletin of Economics and Statistics* 54, 187-207.

[22] Welfe A., Majsterek M., Florczak W., (1994), Model pętli inflacyjnej w gospodarce polskiej - analiza kointegracyjna [The Model of the Inflationary Loop in the Polish Economy- Cointegration Analysis], *Przegląd Statystyczny* 41, 245-264. (in Polish)

[23] Welfe A., Majsterek M., (2000), Analiza kointegracyjna. Sprzężenie inflacyjne [The analysis of cointegration. The inflationary feedback], [in:] *Gospodarka Polski w okresie transformsacji. Zasady modelowania ekonometrycznego*, [ed.:] A. Welfe, PWE, Warszawa, 79-107. (in Polish)

[24] Welfe A., Majsterek M., (2002), Wage and Price Inflation in Poland in the Period of Transition: the Cointegration Analysis, *Economics of Planning* 35, 205-219.

[25] Wróblewska J., (2008), Bayesian Models in the Analysis of Cointegration (with Application to Modeling Inflation in Poland), [in:] *Metody ilościowe w naukach ekonomicznych, Ósme Warsztaty Doktorskie z Zakresu Ekonometrii i Statystyki*, [ed.:] A. Welfe, Oficyna Wydawnicza SGH, Warszawa, 295 - 305.

[26] Zellner A., (1971) *An Introduction to Bayesian Inference in Econometrics*, J. Wiley, New York.

# Appendix - the dataset

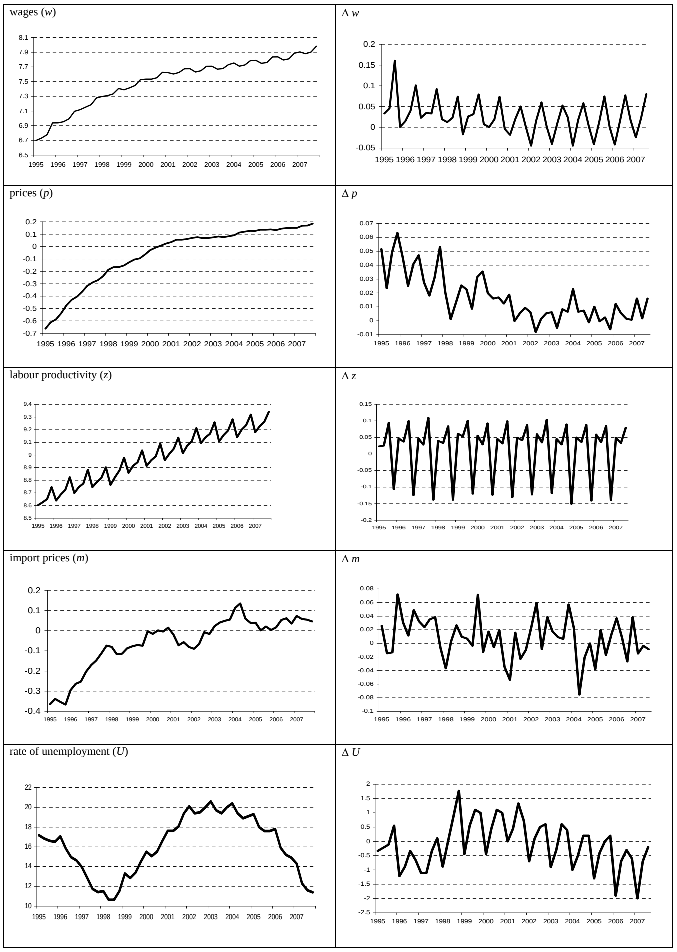


Figure 1: The analysed data: levels and differences